

# AI时代的科学研究

——从知识问答到自主科学发现的范式革命进行时

马滢青（北京大学）

[yqma@pku.edu.cn](mailto:yqma@pku.edu.cn)

第一届中国电子离子对撞机相关物理年会  
青岛，2026/04/21



北京大学



# 大纲

---

## I. 人工智能发展现状

## II. 科研通用智能体——以太

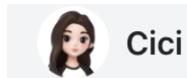
## III. 科研Skills

# AI攻城掠地

结束的  
开始:

2022

- Chat client



2023

- Translation



2024

- Coding



2025

- Mathematics

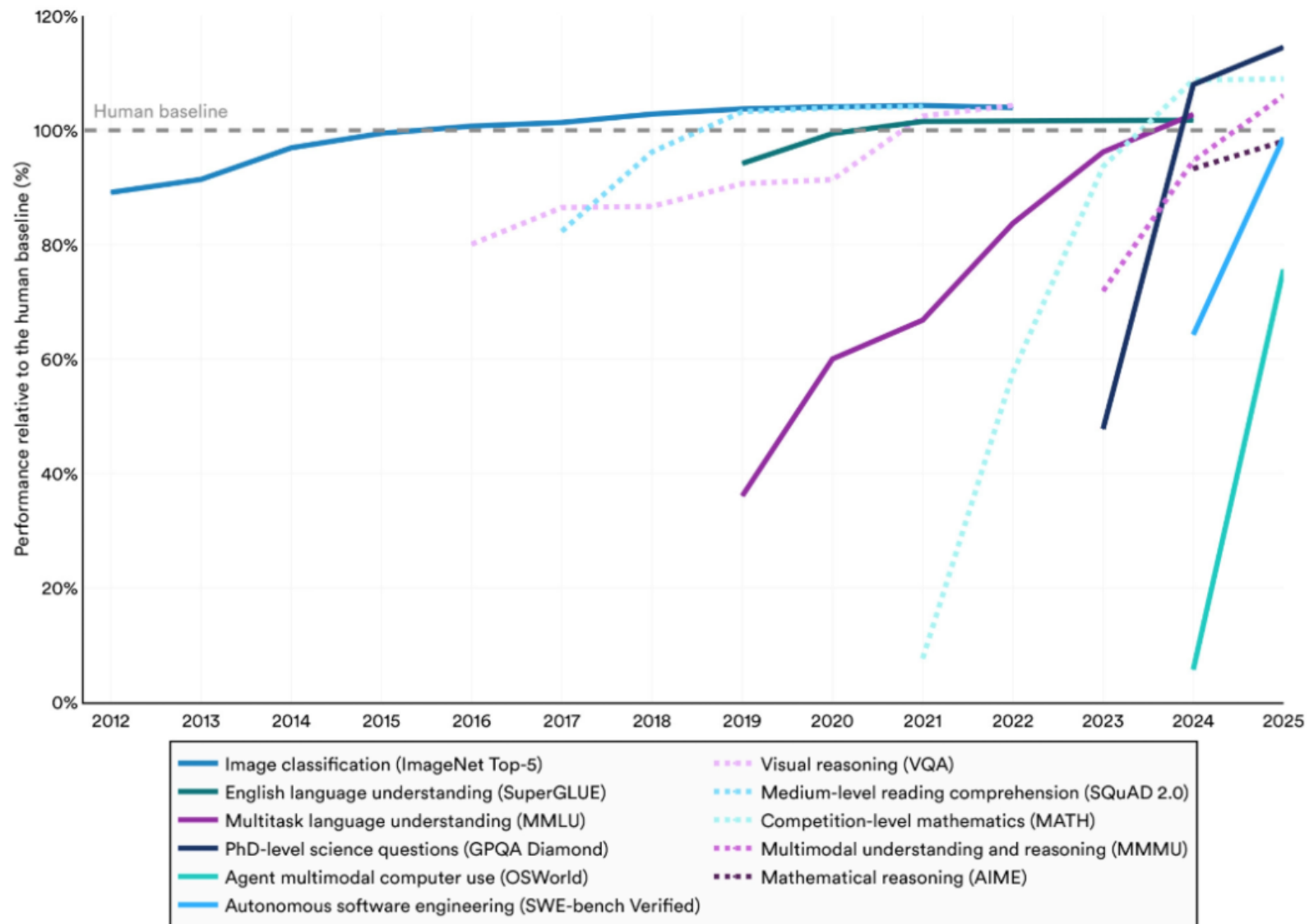


SkyDiscover

# AI与人类能力大对比

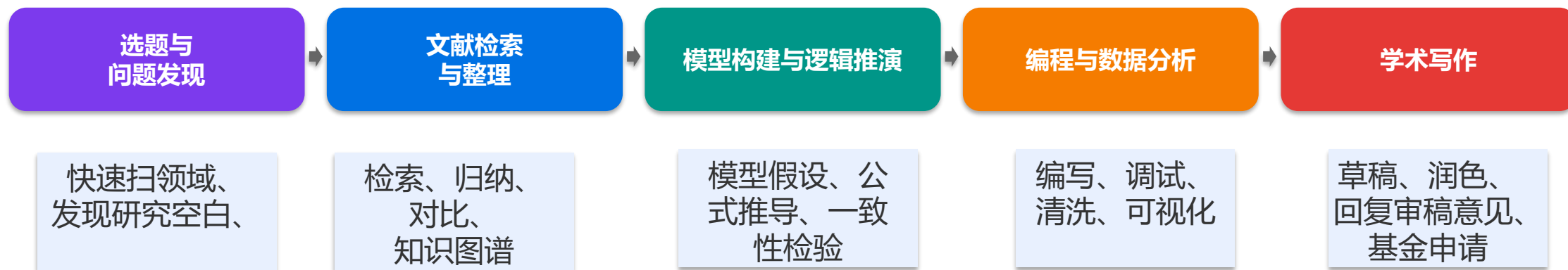
Select AI Index technical performance benchmarks vs. human performance

Source: AI Index, 2026 | Chart: 2026 AI Index report



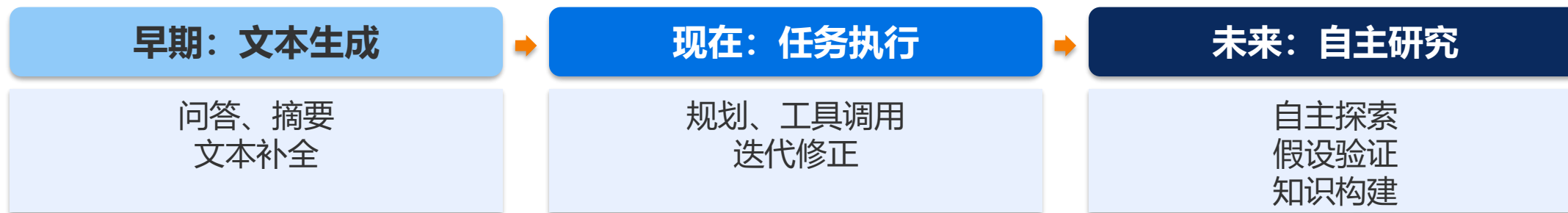
# AI赋能科研

➤ AI已经真正进入科研环节，科研工作流程正在被重新定义



问题已经不是要不要用AI，而是如何更好地使用AI

# 从"会说"走向"会做": 能力边界的变化



未来已在路上

# 大纲

---

I. 人工智能发展现状

II. 科研通用智能体——以太

III. 科研Skills

# AI为何进展如此之快

## 大模型能力

理解、推理、工具调用、  
生成能力持续增强  
(OpenAI、Anthropic、  
Google、字节、  
DeepSeek、智谱等)

## 算力与工程

训练与推理效率  
显著提高  
(英伟达、华为等)

四大驱动力  
共同推动AI跃迁

## 数据与多模态

海量数据覆盖  
多模态融合  
(互联网、书籍、文献、  
用户输入等)

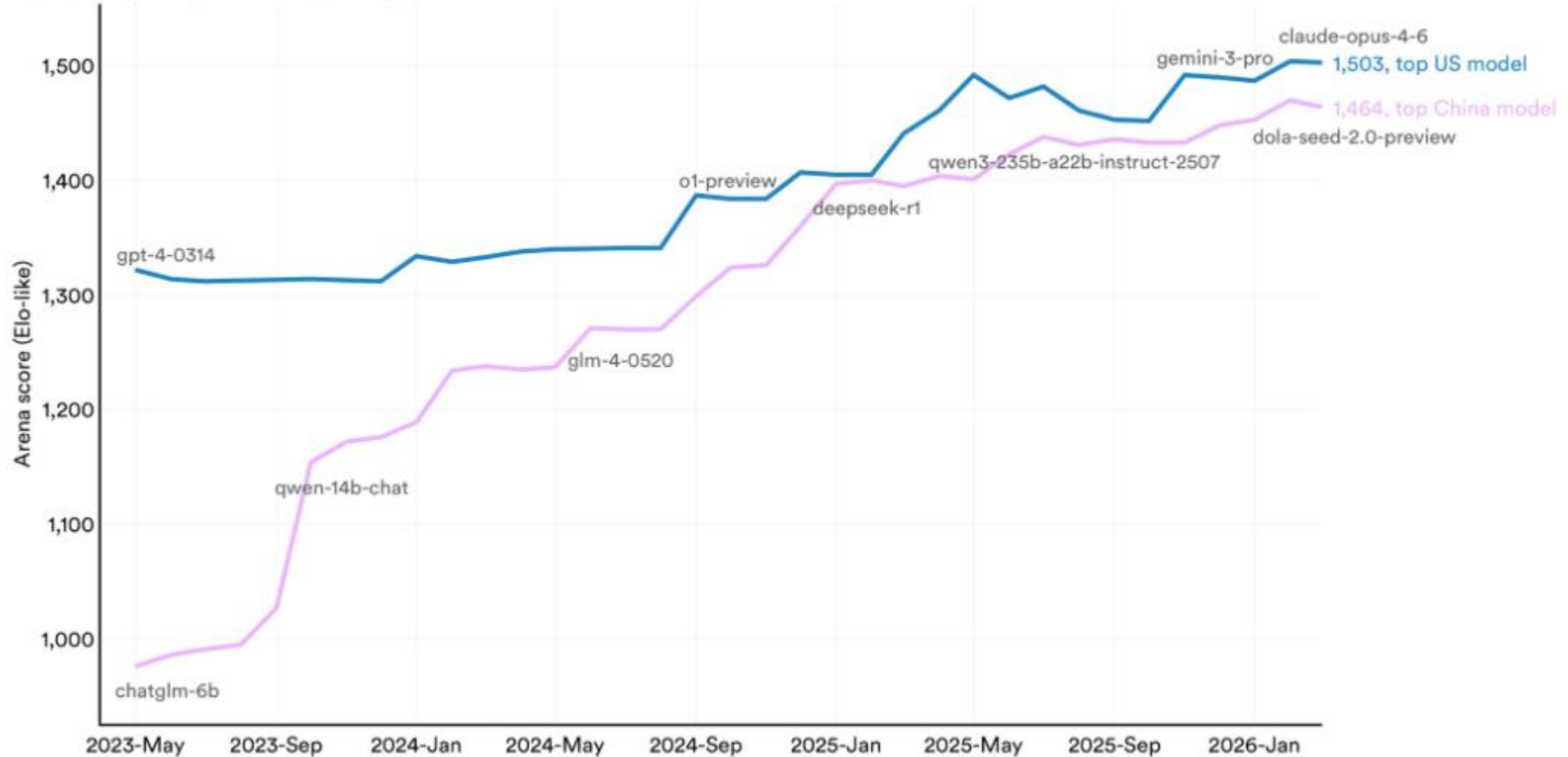
## 工具链建设

Agents、skills、tools  
快速发展

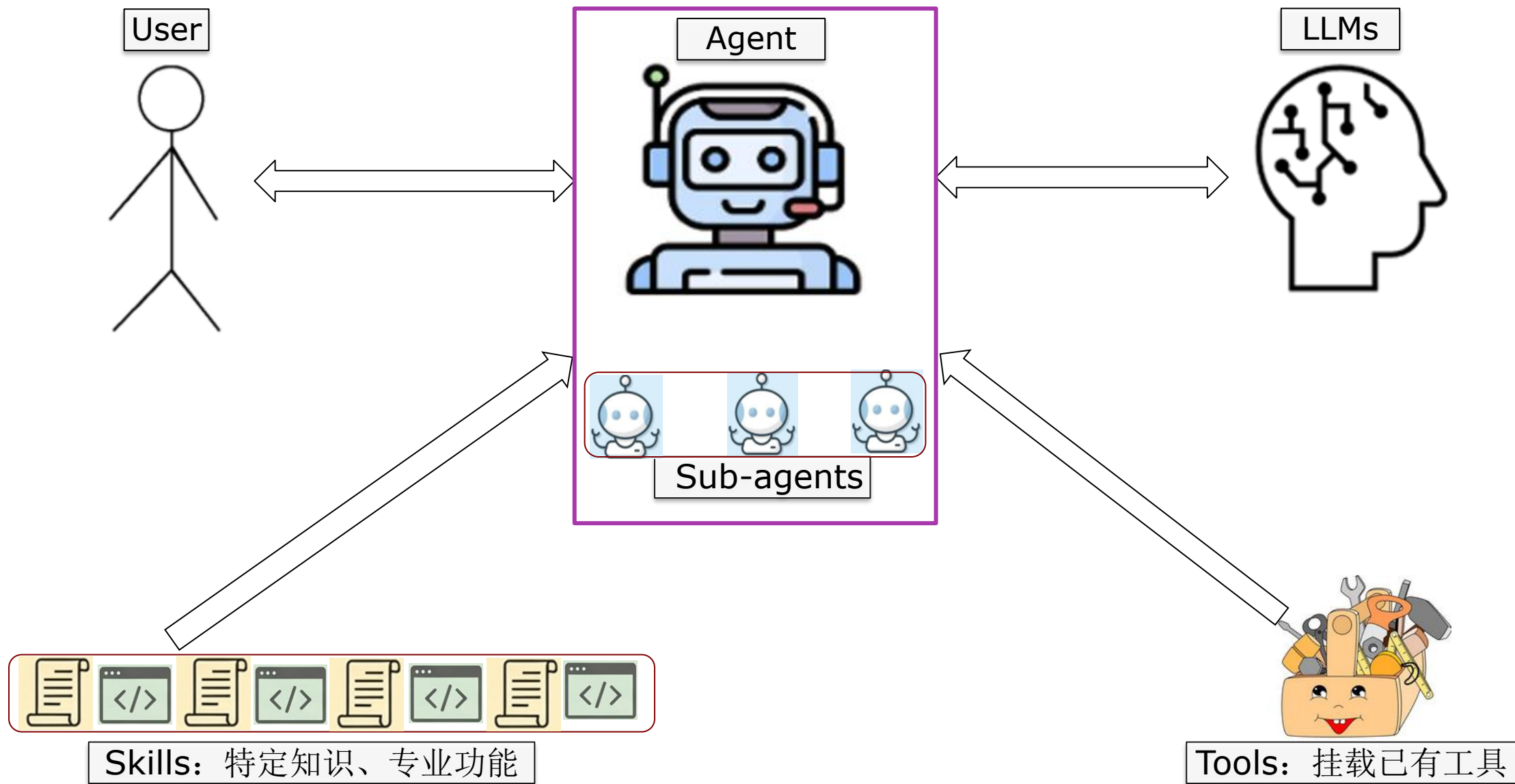
# 中美大模型能力大对比

## Performance of top United States vs. Chinese models on the Arena

Source: Arena, 2026 | Chart: 2026 AI Index report



# AI工具链示意图



# 通用智能体现状

➤ 全部面向编程、国外闭源为主

智能体	年份	开源	流行度
Cursor	2023	否	高
Cline	2024	是	中
Claude Code	2025	否	高
Codex	2025	否	高
<b>OpenCode</b>	<b>2025</b>	<b>是</b>	<b>高</b>

需要科研专属的智能体：

- 编程不可或缺分，但科研远不止编程
- 当前AI编程智能体并非为科研工作流设计
- 科学研究不应受制于闭源工具

Github 146K星

# Aether (以太) 通用智能体

➤ Autonomous Engine for Theoretical & Hands-on Exploration in Research

➤ 2026.3.25发布0.1.0版，4.21发布0.5.2版

- 基于OpenCode开发，打造国际顶级教学科研AI智能体
- 极为便利的安装、使用
- 已有数百位科研人员小范围试用

<https://aether.aiphys.cn/>

## 按平台下载

按实际下载成功事件统计平台分布。

平台	下载次数
Windows	751
macOS	226
Linux	130

## 按版本下载

仅统计能解析到版本号的实际下载成功事件。

版本	下载次数
0.5.1	540
0.4.0	143
0.3.0	140

后台统计的下载次数 12/27

# 以太使用案例：计算前的澄清

## Step 1

### First clarifies scope and checks the environment.

- Interprets the task before launching tools.
- Finds a configuration issue before heavy computation.
- Makes assumptions explicit and asks targeted clarifications.

Careful collaborator,  
not a black-box  
executor.

两圈四点doublebox主积分在s=1,t=10展开计算

请帮我计算两圈四点doublebox拓扑的主积分在运动学量s=1, t=10下的结果，展开到维数正规化因子epsilon的正1阶。

已探索 4 次读取, 1 次搜索

Shell Checks AMFlow environment status

已探索 1 次搜索

我先检查了本机环境：Mathematica 和 AMFlow 都能找到，但当前 Blade 的路径配置需要先修一下，否则没法真正跑两圈积分。

我这里有一个关键歧义需要你确认，我建议默认这样做：

- 按“无质量平面四点两圈 double box”标准族来算；
- 外腿满足  $p_1^2=p_2^2=p_3^2=p_4^2=0$ ,  $(p_1+p_2)^2=s=1$ ,  $(p_2+p_3)^2=t=10$ ;
- 结果采用 AMFlow 的归一化：每圈  $1/(i*\text{Pi}^{(D/2)})$ ;
- 并且我自动修复本机 Blade 配置后再计算。

如果你同意这个默认设定，我就直接继续算到  $\text{epsilon}^1$ ；如果不是这个族/归一化，请直接给我：

- 传播子列表；
- 你说的“主积分”具体是哪一组；
- 归一化约定。

# 以太使用案例：基于任务描述执行

## Step 2

**After the scope is fixed, Aether executes the toolchain and returns a reproducible result.**

- Computes the requested integral and epsilon expansion.
- Returns result, script path, output path, logs, and runtime.
- Preserves provenance for inspection, rerun, and reuse.

**Results come with files, logs, and execution traces.**

按你确认的无质量平面两圈四点 `double box` 标量主积分 `j[dbox,1,1,1,1,1,1,0,0]` 在 `s=1, t=10`、AMFlow 归一化 `1/(i*Pi^(D/2))` 每圈下，展开到 `epsilon^1` 的结果是

```
I(s=1,t=10) =  
- 0.36363636363636363636363636363636/eps^4  
+ (1.50974560756401050956 - 0.85679799643357997413 i)/eps^3  
- (0.3029311826255851266 - 5.0981381811778132198 i)/eps^2  
- (10.242754084619124486 + 9.901633151238733831 i)/eps  
+ (34.351002915426986934 - 20.708845586141352035 i)  
+ (48.059272314275467673 + 78.808210752995366071 i) eps  
+ 0(eps^2)
```

补充信息：

- 计算脚本：`amflow_runs/doublebox_s1_t10/run_amflow.wl`
- 结果文件：`amflow_runs/doublebox_s1_t10/result`
- 日志文件：`amflow_runs/doublebox_s1_t10/.amflow_log`
- 成本估计给出 17 个 `eps` 采样点，实际 AMFlow 完成时间约 510s，全流程约 708s
- 运行结束时 `WARNINGS_DETECTED=0`

如果你愿意，我下一步可以：

1. 把这个结果整理成 Mathematica / LaTeX 可直接粘贴的格式
2. 换成你更常用的 double-box 传播子约定再重写一版结果
3. 继续算同一家族的其他主积分或做更高阶 `epsilon` 展开

# 以太使用案例：交互式任务细化

## Step 3

**When the user corrects the scope, Aether adapts the workflow without restarting from scratch.**

- Updates the target from one integral to the full family.
- Uses existing context instead of re-asking everything.
- Keeps the dialogue continuous while expanding the task.

**Human correction stays inside the workflow.**

不好意思我之前给你的任务可能又歧义，你应该算同一家族的所有主积分的结果，而非这一个积分的结果

对，你这个更准确。我刚才已经改成“同一家族的全部主积分”来算了。

这个质量为零的平面两圈四点 `double box` 家族，在你给定的传播子约定下，Blade/AMFlow 在扇区映射后识别出 8 个主积分，见 [amflow\\_runs/doublebox\\_s1\\_t10/cache/dbox\\_amflow/0/cache/dbox\\_sector mappings/results/rawmasters](#)

归一化仍是每圈  $1/(i*\pi^{(D/2)})$ ，运动学点  $s=1$ ,  $t=10$ ，展开到  $\text{eps}^1$ :

```
j[dbox,1,1,1,1,1,1,1,0,-1] =  
- 2.38636363636363636363636363636363/eps^4  
+ (8.2046549297536126167 - 7.8539816339744830962 i)/eps^3  
+ (14.5887699377630140698 + 39.1997231543622337782 i)/eps^2  
- (104.5125689302236810173 - 21.9839348052435963715 i)/eps  
+ (-47.2313832126498520634 - 210.9974566493930405956 i)  
+ (357.6803364936119324183 - 121.0745302325698511929 i) eps  
  
j[dbox,1,1,1,1,1,1,1,0,0] =  
- 0.36363636363636363636363636363636/eps^4  
+ (1.50974560756401050956 - 0.85679799643357997413 i)/eps^3  
- (0.3029311826255851266 - 5.0981381811778132198 i)/eps^2  
- (10.2427540846191244859 + 9.901633151238733831 i)/eps  
+ (34.3510029154269869336 - 20.7088455861413520352 i)  
+ (48.0592723142754676727 + 78.8082107529953660709 i) eps
```

# 以太使用案例：移动端同步协作（微信、飞书）

微扰量子场论最新进展LaTeX撰写

给我调研一下微扰量子场论最新的进展，并写成 latex文件

Deep-Research

我将从多个角度研究微扰量子场论的最新进展。

Webfetch <https://arxiv.org/search/?query=perturba...>

Webfetch Request failed with status code

Webfetch <https://arxiv.org/search/?query=perturba...>

Webfetch <https://arxiv.org/search/?query=scatterin...>

Webfetch <https://arxiv.org/search/?query=celestial...>

Webfetch <https://arxiv.org/search/?query=resurgen...>

Webfetch <https://arxiv.org/search/?query=amplitud...>

Webfetch <https://arxiv.org/search/?query=BV+qua...>

Webfetch <https://arxiv.org/h/?query=post-Mi...>

以太ClawBot AI

test · 微扰量子场论最新进展LaTeX撰写 · Build · myprovider/glm-5.1

可用命令:

- /n, /new 开启新对话
- /stop 停止当前执行
- /c, /compact 压缩当前上下文
- /m, /model 查看可用模型
- /m l 查看全部模型
- /m n 切换编号模型
- /a, /agent 查看当前模式
- /a <name> 切换指定模式
- /p, /project 查看最近项目
- /p l 查看全部项目
- /p n 切换编号项目
- /p <path> 切换到指定路径
- /s, /session 查看最近会话
- /s l 查看全部会话
- /s n 切换编号会话
- /h, /help 显示帮助信息
- /help list 显示全部命令

/h

test · 微扰量子场论最新进展LaTeX撰写 · Build · myprovider/glm-5.1

会话列表:

- 微扰量子场论最新进展LaTeX撰写 ◀

2026-04-19 19:16

💡 /s n 切换会话 | /s l 查看全部

研究进展如何了

test · 微扰量子场论最新进展LaTeX撰写 · Build · myprovider/glm-5.1

已完成。文件 E:  
`\test\perturbative_qft_survey.tex` 已写入，包含8个章节、30篇参考文献，覆盖复苏理论、天体全息、后闵可夫斯基展开、BV-BFV量子化、振幅自举、变分方法、弯曲时空微扰QFT、约束瞬子等方向。编译命令：  
`xelatex perturbative_qft_survey.tex`

帮我编译成 pdf 文件


test · 微扰量子场论最新进展LaTeX撰写 · Build · myprovider/glm-5.1

编译成功。PDF已生成: E:  
`\test\perturbative_qft_survey.pdf`, 共11页。

发给我

perturbative\_qft\_survey.pdf

205.0 KB

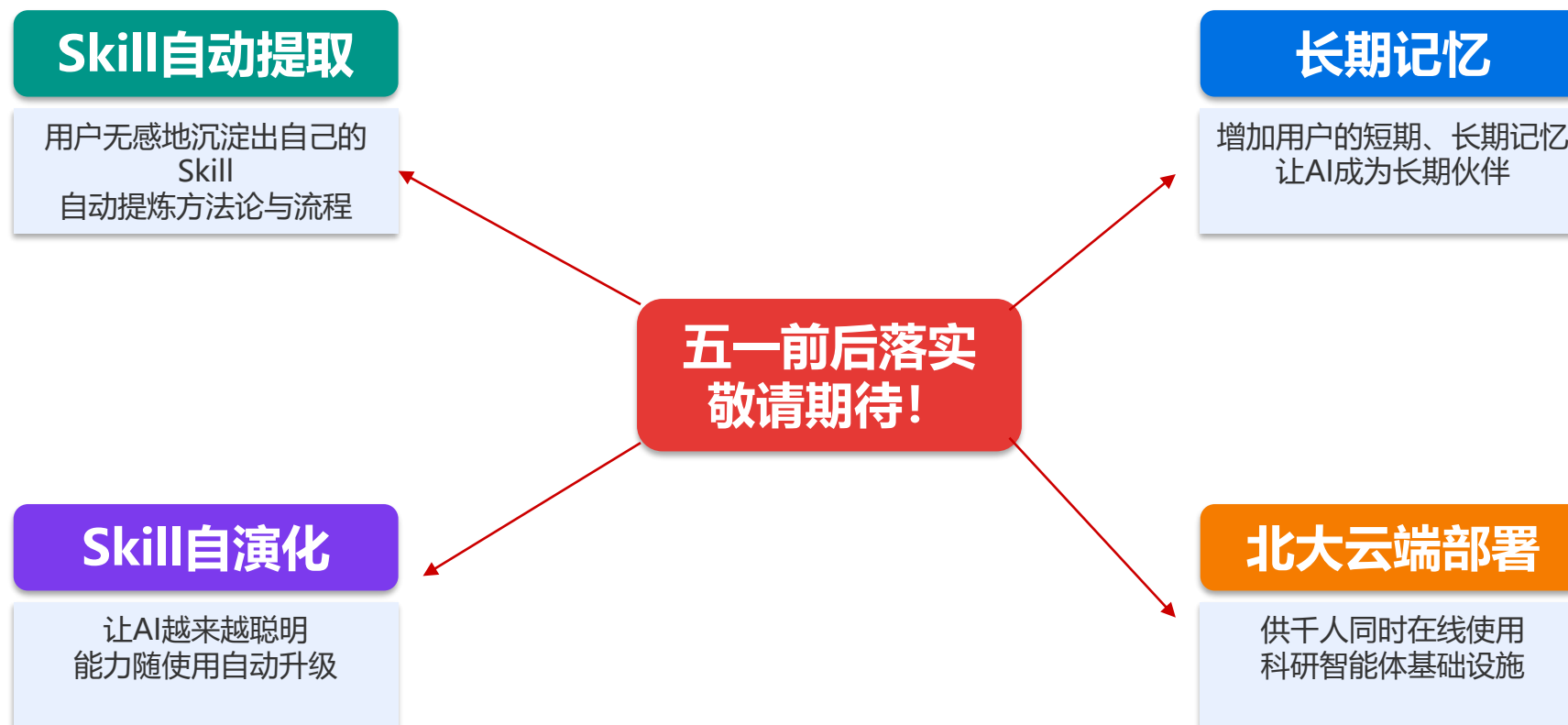


test · 微扰量子场论最新进展LaTeX撰写 · Build · myprovider/glm-5.1

E:\test\perturbative\_qft\_survey.pdf

# 以太智能体展望

## ➤ 让AI成为懂你的科研伙伴



# 大纲

---

I. 人工智能发展现状

II. 科研通用智能体——以太

**III. 科研Skills**

# Skill 的重要性

## ➤ 知识打包

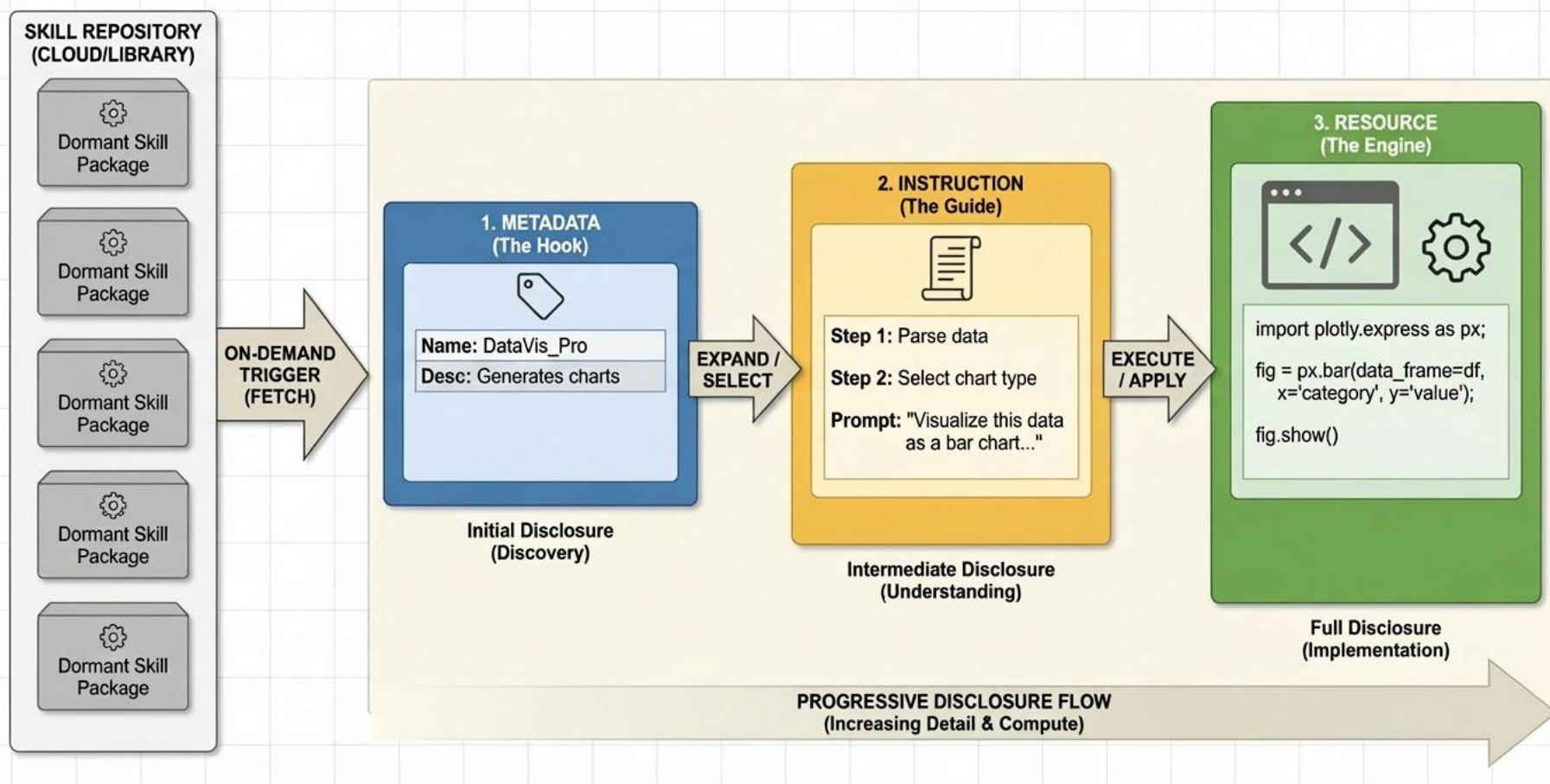
- 领域文本知识
- 特定工具调用
- 专业功能嵌入

## ➤ 渐进式披露

- 按需加载
- 节省上下文
- 不同任务互不干扰

## ➤ 灵活使用

- 易编写
- 易测试
- 易分享



# Skills构建流程

## ➤ 先让Agent完成一个任务

@data/f1\_data.csv 这是 $\Lambda \rightarrow p\ell\nu_\ell$ 半轻衰变形状因子 $f_1$ 的数据, 进行拟合。  
所有文件放在fit-without-skills/test-qwen中

已探索 1 次读取

Shell Count lines in f1\_data.csv

Shell Preview first 20 columns of CSV

Shell Count columns in CSV

Shell Check end of CSV file format

Brainstorming

我已了解数据格式:

- 6行 × 500列, 每行是一个配置 (config), 每列是一个数据点
- 数据值范围约 -0.87 到 -1.35, 呈递减趋势 (行索引越大, 值越负)

这是  $\Lambda \rightarrow p\ell\nu_\ell$  半轻衰变形状因子  $f_1$  的格点QCD数据。在开始拟合之前, 我需要确认几个问题:

第1问: 物理变量是什么?

这500个数据点对应的自变量是什么? 通常形状因子拟合是关于  $q^2$  (动量转移平方) 的函数。请问:

- A. 每列对应不同的  $q^2$  值 (需要另外提供  $q^2$  数组)
- B. 行索引代表不同的  $a^2$  (格点间距平方) 或  $m_\pi^2$  ( $\pi$ 介子质量平方), 需要进行手征/连续极限外推
- C. 您有单独的  $q^2$  或其他自变量数据文件
- D. 其他情况 (请说明)

Plan · qwen3.6-plus · 33秒

1. 每行对应不同的 $q^2$ 值。从上到下-0.05, -0.04, -0.03, -0.02, -0.01, 0

明白了! 6个  $q^2$  点, 每行500个bootstrap样本: \*\*平均每个500个样本以获取中心值和协方差矩阵, 然后进行拟合。现在需要确认拟合方案:

第2问: 拟合函数形式?

对于  $\Lambda \rightarrow p\ell\nu_\ell$  半轻衰变形状因子  $f_1(q^2)$ , 常用的参数化有:

- A. z-展开 (z-expansion, BKGL/BSZ形式, 推荐) —— 基于解析性和么正性, lattice QCD标准做法
- B. 偶极形式 (Dipole:  $f_1(q^2) = \frac{f_1(0)}{(1-q^2/M^2)^2}$ ) —— 简单, 常用
- C. 多项式 ( $f_1(q^2) = a_0 + a_1q^2 + a_2q^4 + \dots$ ) —— 简单但外推不稳定
- D. 单极点形式 (Monopole:  $f_1(q^2) = \frac{f_1(0)}{1-q^2/M^2}$ )

第3问: 输出目录确认

所有文件放在 `fit-without-skills/test-qwen` 中 (会创建该目录), 可以吗?

# Skills构建流程

## ➤ 先让Agent完成一个任务

拟合函数形式采取A和B，输出目录为fit-without-skills/test-qwen  
执行

Shell Create output directory

写入 fit\_f1.py /fit-without-skills/test-qwen/

已探索 1 次读取

I need to fix the undefined `OUTPUT_DIR` variable.

编辑 fit\_f1.py /fit-without-skills/test-qwen/ +14 -1

Now let me run the fitting script:

Shell Run the form factor fitting script

❌ edit 失败

已探索 2 次读取

编辑 fit\_f1.py /fit-without-skills/test-qwen/ +1 -1

Shell Run the form factor fitting script

拟合已完成，所有输出文件已生成到 `fit-without-skills/test-qwen/`。以下是核心结果：

### 拟合结果摘要

模型	$\chi^2/\text{ndof}$	$f_1(0)$	斜率 $f_1'(0)$
z-exp(N=1)	1.81/5 (p=0.87)	-1.109 ± 0.071	-1.395 ± 0.089
z-exp(N=2)	0.03/4 (p=1.00)	-1.168 ± 0.074	-3.766 ± 0.260
dipole	0.02/4 (p=1.00)	-1.170 ± 0.075	-4.001 ± 0.281
monopole	0.01/4 (p=1.00)	-1.171 ± 0.075	-4.181 ± 0.296

### 输出文件：

- `results.txt` — 完整拟合结果
- `covariance_matrix.csv`, `correlation_matrix.csv` — 协方差/相关矩阵
- `mean_f1.csv` — 中心值与误差
- `fit_plot.pdf/png` — 拟合曲线图
- `residual_plot.pdf/png` — 残差图
- `*_distributions.pdf/png` — 参数 bootstrap 分布

# Skills构建流程

## ➤ 根据执行流程，调用**skill-creator**

总结上述拟合流程，生成一个skill

Skill-Creator

Shell Initialize lattice form factor fit skill

Shell Initialize lattice form factor fit skill

已探索 3 次读取

写入 fit\_formfactor.py /fit-without-skills/lattice-formfactor-fit/scripts/

写入 SKILL.md /fit-without-skills/lattice-formfactor-fit/ >

Shell Test the skill's fit script

Shell Test the skill's fit script with python3

Shell Package the lattice-formfactor-fit skill

已探索 2 次读取

Skill 已生成完毕，保存在 [fit-without-skills/lattice-formfactor-fit/](#)。

# Skill使用效果

- 用户提出问题，触发**skill**:

```
@data/f1_data.csv 这是 $\Lambda$  to  $p$   $\nu$   $\ell$   $\ell$   $\ell$  半轻衰变形状因子f1的数据, 进行拟合。
所有文件放在fit-with-skills/test-glm中

Build · glm-5.1 · 2026-04-19-14:47
```

Semileptonic

已探索 1 次读取

数据已加载，共有6行（对应6个 $q^2$ 点），每行包含大量bootstrap样本值。我需要确认一些关键信息：

- 针对性与人类用户互动:

数据已加载，共有6行（对应6个 $q^2$ 点），每行包含大量bootstrap样本值。我需要确认一些关键信息：

问题 2 已回答

这6行数据对应的 $q^2$ 值分别是什么？请提供6个 $q^2$ 值（例如：-0.4, -0.2, 0, 0.2, 0.4, 0.6）  
-0.05 to 0, space 0.01

这是bootstrap数据吗？即每行对应一个 $q^2$ 点，每列是一个bootstrap样本？  
是bootstrap数据

- 自动调用工具:

已有现成的拟合脚本。需要更新路径（sys.path指向当前skill位置），创建output目录，然后运行。

编辑 fit\_f1.py /fit-with-skills/test-glm/ +134 -41

Shell Create output directory

```
$ mkdir -p /Users/zhangsan/Desktop/presentation/fit-with-skills/test-glm/output
```

# Skill使用效果

- 根据具体情况优化/重写  
工具或流程:

The f-string is interpreting `ndf` as a variable inside the markdown. Need to fix the f-string to avoid this issue.

已探索 1 次读取

问题在于 `\mathrm{ndf}` 在 Python f-string 中被当作转义序列, 且 `ndf` 被当作 Python 变量。我来彻底重写一下结果字符串部分:

编辑 `fit_f1.py` /fit-with-skills/test-glm/ +66 -18

- 结构化输出结果:

拟合完成。结果如下:

方案	$f_0$	$a_1$	$\chi^2/\text{ndf}$
Dipole	$-1.1737 \pm 0.0746$	$3.295 \pm 0.115$	0.0042
z-Series	$-1.1714 \pm 0.0744$	$-32.90 \pm 1.24$	0.0074

两种方案的  $\chi^2/\text{ndf}$  都远小于1 (数据点只有6个而参数有2个), 拟合质量良好。所有输出文件已保存在 `fit-with-skills/test-glm/output/` 目录中。

# 测试与优化Skills

## ➤ 使用编写好的**skill**处理具体任务，进一步优化**skill**能力

- 以表格形式总结

### Form Factor Fit Results

Decay:  $\Lambda \rightarrow p \ell \nu_\ell$  Kinematics:  $M_A = 1.115683$  GeV ( $\Lambda$ ),  $M_B = 0.938272$  GeV ( $p$ ),  $m_{pole} = 0.89166$  GeV ( $K^*(892)$ )  
 $q^2$  values: -0.05, -0.04, -0.03, -0.02, -0.01, 0.0 \*\* $q_{max}^2 = 0.031475$  GeV<sup>2</sup>

f1

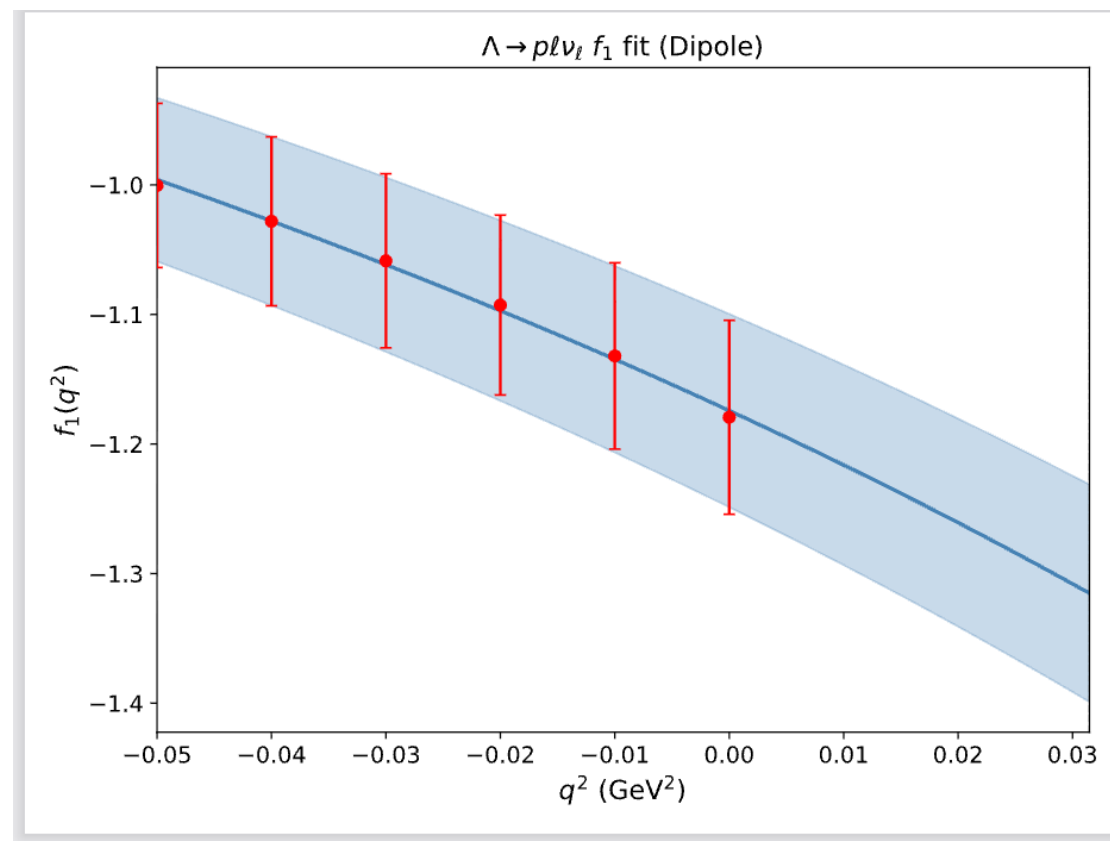
Dipole

Parameter	Value	Error
$f_0$	-1.174225	0.074595
$a_1$	3.322115	0.115853

z-Series

Parameter	Value	Error
$f_0$	-1.172198	0.074463
$a_1$	-33.292375	1.249125

- 以图像形式总结



# 分享与获取skills

- 获取他人分享的**skills**
- 上传自己的**skills**
- 计划：由科学共同体审核质量

The screenshot displays the Skill Market interface. At the top, it says "SKILL MARKET" and "发现下一项关键能力" (Discover your next key skill). The URL "skill.aiphys.cn" is visible. Below the header is a search bar and filters for "全部分类" (All categories), "最新更新" (Latest updates), and "全部状态" (All status). Three skill cards are shown:

- amflow-compute** (ID: amflow-compute-v2): "已发布" (Published). Description: "amflow-compute skill 是一个用于\*\*自动化计算 Feynman 圈积分和相空间积分\*\*的工作流技...". Tags: quantum field theory, perturbative, Feynman integral, multi-loop. Author: 李想pku, 67 likes, 26 days ago. "查看详情" (View details) link.
- paper\_summarizer** (ID: paper-summarizer): "已发布" (Published). Description: "主要用于把英语论文翻译概括为短小精悍格式清晰条理分明的中文论文, 将正文和图片表格分...". Tag: 论文格式化与编辑, 论文翻译, 加速文献阅读. Author: 吴弈笛, 31 likes, 20 days ago. "查看详情" (View details) link.
- AICodingSkill** (ID: aicodingskill): "已发布" (Published). Description: "主要用在当你有了一个程序的大致想法后用AI写出方便使用清晰易懂的代码的同时人类参与者...". Tag: 程序设计, 加强人机交互. Author: 吴弈笛, 20 likes, 20 days ago. "查看详情" (View details) link.

# 我可以做什么?

## ➤ 从"试用AI"升级为"重构科研流程"

### 个人研究者 及课题组

从一个高频任务开始, 建立AI workflow, 学会验证而非盲信  
沉淀可复用的skills, 构建起课题组研究助手

善用AI: 让AI处理低效重复环节

验证AI: 关键结论必须回到原始验证

驾驭AI: 人类保留问题定义与最终判断权

未来高水平科研的重要能力之一, 是与AI高效协作

# *Thank you!*